

Language Documentation and Description

ISSN 1740-6234

This article appears in: *Language Documentation and Description, vol 10: Special Issue on Humanities of the lesser-known: New directions in the description, documentation and typology of endangered languages and musics*.
Editors: Niclas Burenhult, Arthur Holmer, Anastasia Karlsson, Håkan Lundström & Jan-Olof Svantesson

Elusive articles in Sulawesi: between syntax and prosody

RENÉ VAN DEN BERG

Cite this article: René van den Berg (2012). Elusive articles in Sulawesi: between syntax and prosody. In Niclas Burenhult, Arthur Holmer, Anastasia Karlsson, Håkan Lundström & Jan-Olof Svantesson (eds) *Language Documentation and Description, vol 10: Special Issue on Humanities of the lesser-known: New directions in the description, documentation and typology of endangered languages and musics*. London: SOAS. pp. 208-227

Link to this article: <http://www.elpublishing.org/PID/119>

This electronic version first published: July 2014



This article is published under a Creative Commons License CC-BY-NC (Attribution-NonCommercial). The licence permits users to use, reproduce, disseminate or display the article provided that the author is attributed as the original creator and that the reuse is restricted to non-commercial purposes i.e. research or educational use. See <http://creativecommons.org/licenses/by-nc/4.0/>

EL Publishing

For more EL Publishing articles and services:

Website:	http://www.elpublishing.org
Terms of use:	http://www.elpublishing.org/terms
Submissions:	http://www.elpublishing.org/submissions

Elusive articles in Sulawesi: between syntax and prosody

René van den Berg

1. Introduction

Articles are common in the world's languages, as can be seen in Chapters 37 and 38 of the *World atlas of language structures* (WALS) (Dryer 2008).¹ They usually code definiteness (e.g. Dutch *de* and *het*) or indefiniteness (e.g. Dutch *een*), but sometimes specificity or referentiality. Article-like elements also occur in the 1,200+ languages of the Austronesian family (spoken in insular Southeast Asia and most of the Pacific), but their function is usually very different from that of the articles of Indo-European languages. Rarely do they indicate definiteness or specificity as such, although in Philippine languages a morpheme that appears to be an article often uniquely marks the 'subject' (or topic), which is usually definite. In some Austronesian languages (especially in Oceania) such morphemes appear to be little else than simply noun markers or 'signs of nominality', adding little or no meaning to the clause apart from making the word class explicit. The typologically somewhat unexpected behaviour of these morphemes in Austronesian has also led to considerable terminological variation, especially in the Philippines, where these NP introducing morphemes have been classified as case markers, articles, determiners, phrase markers, case particles, construction markers and much more (see Reid 2002 for a survey). WALS also excludes these elements from its survey of articles. Whatever their function, and whatever the best terminology is to describe them, it is clear that these morphemes play a big role in Austronesian. In fact, Proto-Austronesian is reconstructed with an extended set of these 'case markers' (Ross 2006, Blust 2009: 444).

In Sulawesi (eastern Indonesia) we find a number of Austronesian languages which have noun-introducing morphemes which are hard to describe, as they perform unclear or unexpected functions, even from an

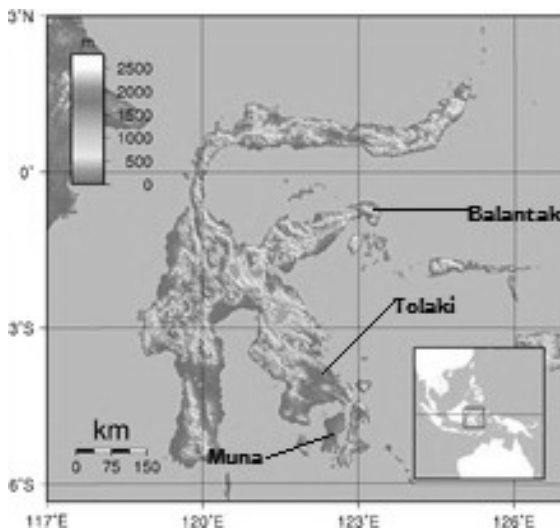
¹ An earlier version of this article was presented at the conference Humanities of the Lesser Known in Lund (Sweden), September 2010. I would like to express my thanks to the audience for helpful feedback, and to David Mead and Scott Youngman for giving me access to unpublished Tolaki material. Thanks also to Paul Kroeger for taking a critical look at this article and an anonymous reviewer for suggesting an alternative approach to the Balantak data. The Tolaki data presented here does not cover all the facts relating to the Tolaki article; several complications remain unmentioned.

Austronesian perspective. This paper looks at such elusive morphemes in three languages of Sulawesi: Balantak *a*, Muna *o* and Tolaki *o*=, with the latter two displaying an unusual interaction between syntactic and prosodic features. I simply refer to these morphemes as articles for lack of a better cover term. The languages are part of the large Celebic group (Mead 2003). The form of the article is cognate, but the languages themselves belong to three different microgroups within Celebic and are not closely related. Typologically they differ considerably in their core grammatical system of voice marking, grammatical alignment and verbal inflection, but they also share a number of distinctive characteristics, including the following:

- five vowels, penultimate stress, few consonant clusters, a tendency for open syllables;
- disyllabic content words;
- rich derivational morphology on the verb;
- no gender distinction, but inclusive and exclusive first person pronouns;
- a rich set of demonstratives.

This paper is organised as follows. In sections 2, 3 and 4, I present the data for each of the three languages under discussion. Section 5 provides a summary, section 6 a diachronic perspective, while section 7 concludes with a number of remaining questions.

Figure 1: Map of Sulawesi



2. Balantak

Balantak is spoken by about 30,000 people at the tip of the eastern arm of Sulawesi (see Figure 1). It belongs to the Saluan-Banggai subgroup of Celebic. The data for the following section is taken from van den Berg & Busenitz (to appear). Basic constituent order is SVO. The article *a* frequently precedes certain noun phrases. Its main features can be listed as follows:

- *a* only occurs before common nouns, never with proper nouns (which take the article *i*) or pronouns;
- *a* only occurs with subjects which follow the predicate;
- the post-predicate subject marked by *a* can be definite or indefinite, though it is usually definite;
- phonologically, *a* normally cliticises to the preceding word without causing a stress shift, though it is written as a separate word, a convention which I follow in this paper.

Examples, with the article in bold, are the following²:

- (1) (a) *Ma-polos tuu' a sengke'-ku.*
 INTR.IRR-hurt very ART back-1SG
 'My back really hurts.'
- (b) **Ma-polos tuu' sengke'-ku.*
 INTR.IRR-hurt very back-1SG

Preverbal subjects never take the article:

- (2) (a) *Sengke'-ku ma-polos tuu'.*
 back-1SG INTR.IRR-hurt very
 'My back really hurts.'
- (b) **A sengke'-ku ma-polos tuu'.*
 ART back-1SG INTR.IRR-hurt very

² The abbreviations used in this paper are: 1 = first person. 3 = third person. ART = article. AV = agent voice. CAUS = causative. INTR = intransitive. IRR = irrealis mode. LIG = ligature. LOC = locative. NEG = negator. PERF = perfective. PV = patient voice. REA = realis mode. REL = relative marker. SG = singular.

Other examples of post-predicate subjects with the article are:

- (3) *Alayo' tuu' a ili-na.*
 high very ART price-3SG
 'The price is very high.'
- (4) *Taasi' titiu' a lewolot men ni-liu.*
 NEG little ART obstacle REL PV.REA-pass.by
 'The obstacles that (we) experienced were not a few.'
- (5) ... *uar-kon-on-mo a panganon.*
 go.out-CAUS-PV.IRR-PERF ART proposal.gift
 '... the marriage proposal gifts are taken out.'

The basic function of *a* is to mark a backed subject. That is, *a* marks a subject which is found at (or, depending on one's theoretical position, has been moved to) the back of a clause because it is given or known information. Because *a* is an enclitic article with common nouns, it follows that there are a number of contexts where it does not occur for neither prosodic nor syntactic reasons.

In the first place, *a* is missing for prosodic reasons when it fails to find a host.³ This is true in single-word clauses such as answers to content questions ('What is this?'), where there simply is no preceding host for *a* to attach to. For the same reason, *a* appears to be lacking with subjects in SV or SVO clauses, as the subject noun phrase is the initial constituent of the clause.⁴

³ An anonymous reviewer suggested an alternative explanation for the lack of *a* with clause-initial subjects, namely that such constituents are synchronically and/or historically topicalised subjects, which – like various Formosan languages – lack the article. From a diachronic perspective, this is certainly a possibility which deserves further exploration. However, there is no synchronic evidence that Balantak is a V-initial language and that initial subjects are topicalised. In intransitive clauses the order SV is slightly more dominant than VS, while in active transitive clauses the unmarked constituent order is SVO; see van den Berg & Busenitz (to appear) for details. The fact that the article is an enclitic which needs a preceding host is a fairly robust generalisation (but see footnote 4), which provides an adequate synchronic explanation for the current distribution.

⁴ There are admittedly clause-initial elements that can precede a preverbal subject NP, such as conjunctions and adverbs, but these are also excluded as hosts. Either their word class blocks the hosting, or their function (they are not predicates), or possibly the constraint has been broadened and generalised to cover all pre-predicate subjects.

Secondly, *a* is missing for syntactic reasons when it does not introduce an NP which is functioning as a subject. The following cases can be distinguished:

- with objects following the verb in agent voice, as in (6):
- (6) *Tumbe-tumbe-na yaku' mang-ala lemba.*
 RED-first-3SG 1SG AV.IRR-take settling.trough
 'First of all I get a settling trough.'
- with agents in patient voice and locative voice.
 - with nominal possessors in a noun phrase.
 - after prepositions such as *bo* 'for', *tia* 'with' and *na* 'general locative preposition'. In this case the prosodic and syntactic reasons reinforce each other. The clitic position has already been taken by the preposition and the noun phrase does not function as a subject.

Finally, in existential clauses with subjects following *isian* 'there is, there are', the article is lacking for reasons which are not yet entirely clear. Apparently the subject in an existential clause does not have all the properties of a canonical subject. Alternatively, the NP in an existential clause is not a subject at all (Clark 1978).

3. Muna

Muna is spoken by over 300,000 people on the island of Muna and some neighbouring islands (see Figure 1). It is a member of the Muna-Buton subgroup of Celebic. All data on Muna is from van den Berg (1989) and from subsequent fieldwork.

The article *o* in Muna illustrates a truth which other linguists have also found in their work, namely that the smallest morphemes in a language are often the hardest to adequately account for. In my 1989 dissertation I dedicated approximately seven pages to discussing this elusive morpheme *o*, the main features of which can be summarised as follows:

- The proclitic article *o* (written as a separate word) only occurs with common nouns, never with pronouns or names.
- The article has no independent meaning. There is no relation with definiteness, specificity, referentiality, topichood or a grammatical

function such as subject. Native speakers are also at a complete loss to assign any meaning to this element.

- NPs in certain positions require the article *o*:
 - single-word utterances;
 - medial elements in enumerations;
 - preverbal nouns;
 - nominal predicates.

The following examples illustrate these four categories.

a. *Single-word utterances*, such as a response to the question ‘What is this?’

- (7) (a) *o medha* ‘a table’
 (b) *o tonde* ‘a glass’
 (c) *o kalei* ‘bananas’

b. *Enumeration*. In an enumeration of three or more common nouns following a verb, *o* is obligatory for the medial element(s), optional for the first element, and absent for the final element (which follows the preposition *bhe* ‘with, and’):

- (8) *ne-gholi (o) pae, o kenta, o kambulu bhe kalei.*
 3SG.REA-buy ART rice ART fish ART vegetables with banana
 She bought rice, fish, vegetables and bananas.’

c. *Preverbal nouns*.

- (9) *O dahu no-kotou.*
 ART dog 3SG.REA-bark
 ‘A dog barks.’ / ‘Dogs bark.’
- (10) *O kapoluka no-bisara-mo ...*
 ART tortoise 3SG.REA-speak-PERF
 ‘The tortoise said...’

Notice that preverbal common nouns are ungrammatical without *o*:

- (11) (a) **dahu* *no-kotou*
 (b) **kapoluka* *no-bisara-mo*

d. *Nominal predicates.*

- (12) *Ama-ku* *o* *guru.*
 father-1SG ART teacher
 ‘My father is a teacher.’

Based on its distribution in texts and its obligatory positions, my conclusion was that *o* marks ‘relative syntactic freedom’ (van den Berg 1989: 108). This can be reworded as follows: the article *o* initiates a noun phrase which follows a major syntactic boundary, typically a noun phrase at the beginning of a sentence, a clause, or a nominal predicate. The notion of ‘major syntactic boundary’ is admittedly vague, but covers three of the four categories illustrated above. In addition, it may be part of the reason why the article never appears after prepositions such as *we* ‘in, at’ or *bhe* ‘with, and’. Phrases such as *we lambu* ‘in a/the house’ and *bhe mie* ‘with people’ can never be expanded with an article before the noun: **we o lambu*, **bhe o mie*. The syntactic reason could be that within a prepositional phrase the link between the preposition and the following noun phrase is too tight for the article to intervene.

However, there are some further distributional features of the article *o* in Muna which seem to be unrelated to the notion of a syntactic boundary. In the first place, the article cannot be combined with the following two elements within an NP: (1) possessive suffixes, as in **o guru-ku* ‘my teacher’, **o dahu-no* ‘his/her dog’. These phrases are ungrammatical in any environment, whether clause-initial clause-medial or as complete utterances. (2) Prenominal measure phrases (that is, a noun preceded by a bound numeral and a classifier), as e.g., **o tolu-ghulu dahu* (ART three-CLAS dog) ‘three dogs’. The correct phrase is *tolu-ghulu dahu* or *o dahu tolu-ghulu*, both meaning ‘three dogs’.

Secondly, there is considerable idiolectal variation in the application of *o* to other NP functions such as postverbal nouns and dependent nouns. Below are the results presented from a small written test in which four speakers (A, B, C and D) comment on the acceptability of the presence of *o* in a number of clauses taken from a folktale. In the table a plus sign means that *o* is obligatorily present, a minus sign that *o* is obligatorily absent and a plus-minus sign (±) that *o* is optionally present. (A fuller version with more examples can be found in van den Berg 1989: 106–108.)

				A	B	C	D	
<i>Preverbal subject:</i>								
(13)	O	<i>karambau</i>	<i>no-mente.</i>	+	+	+	+	
	ART	buffalo	3SG.REA-amazed					
	‘The buffalo was amazed.’							
<i>Postverbal subject:</i>								
(14)	<i>No-sampu</i>	<i>kaawu</i>	o	<i>ure...</i>	-	±	-	±
	3SG.REA-go.down	after	ART	high.tide				
	‘When the high tide went down...’							
<i>Postverbal object:</i>								
(15)	<i>No-salihi</i>	o	<i>karambau.</i>	-	-	±	±	
	3SG.REA-urge	ART	buffalo					
	‘He urged the buffalo.’							

It is clear that all four speakers agree that *o* is obligatory with preverbal nouns, as in (13), but how can we account for the variation in usage with the postverbal noun phrases in (14) and (15)? Is this simply stylistic or idiomatic variation? This was my original hypothesis, but in written texts *o* is actually rare following verbs. Why then were some native speakers fairly tolerant of this? For my dissertation research I was unable to solve this puzzle and my final paragraph on the topic is worth quoting in full (van den Berg 1989: 108):

Possibly, too, this variation is connected with speech tempo. One informant suggested that in slow, careful speech the use of *o* is more appropriate with postverbal nouns than in normal speech. This points in the direction of ‘intonational binding’ as an explaining mechanism. I was not able to confirm this hypothesis by independent recordings. Clearly, this aspect of Muna grammar still needs much research.

Based on further intermittent exposure to the language over the last 20 years I can now confirm that the native speaker who expressed this intuition was on the right track. I must also admit that in my search to discover patterns I was too tied to the written form of the language and was paying insufficient attention to questions of speech tempo and intonation at the time of my dissertation research. The exercise I did with the four native speakers was a written exercise, and I failed to record them reading the examples with or

without the article. A recording with the article in postverbal position would almost certainly have given me the clues I needed.

The breakthrough in my understanding of *o* occurred when I was listening to a text recorded in 2009 in which a widow relates the events surrounding the death of her husband. Several times I was struck by the presence or absence of the article in post-verbal position. In each case it seemed to be linked to the presence or absence of an intonational pause. The following two examples from that text are illustrative. Example (16) illustrates what is most commonly found in the language, that is, the absence of the article with an object noun phrase following the verb (sound file Berg1.mp3).

- (16) *Amba-no* “*A-s<um>uli* *ao-maa* *kaago*.”
 word-3SG 1SG-IRR-return 1SG-IRR.eat medicine
 ‘He said, “I’ll go back to take medicine”.’

The words are spoken as a continuum without a clear pause between the verb *ao-maa* ‘I will eat’ and the object *kaago* ‘medicine’. In example (17), on the other hand, there is a distinct pause following the verb *ne-ala* ‘he took’ (sound file Berg2.mp3).

- (17) *ne-ala* [0.6] *o* [0.4] *paeasa* *bhe* *kadangkuti*
 3SG.REA-take ART mirror with tweezers
 ‘He took a mirror and tweezers...’

It appears then, that the brief hesitation which followed the verb in (17) set off the trigger for the use of the article. In other words, even though there is no major syntactic boundary between the verb and the following object (which form a closely-knit syntactic phrase), a hesitation or a pause creates a prosodic boundary which in turn triggers *o*. Interestingly, *o* itself is followed by another brief pause, but the second pause does not trigger another *o*.

My revised conclusion regarding Muna *o* is therefore as follows. The article *o* is a left-edge boundary marker which is sensitive to both syntactic and prosodic structure. It initiates a noun phrase which follows a major syntactic boundary, but additionally the article *o* also initiates a noun phrase which follows a major prosodic phrase boundary. These are typically nouns in isolation, nouns occurring utterance-initially, nouns in enumerations after pauses and nouns after hesitations, irrespective of the syntactic structure. Notice that prosody is not the only factor. A nominal predicate involving a common noun as illustrated in (12) must be introduced by *o*, even though the whole clause is pronounced as one intonation group, without pause and

without hesitation. It is this combination of syntactic and prosodic factors which makes the description of Muna *o* so elusive and challenging.

Furthermore, as mentioned before, the distribution of *o* is also subject to the following three constraints which also demand an explanation:

- it does not co-occur with possessive suffixes;
- it does not co-occur with bound numerals in a measure phrase;
- it never follows a preposition.

The absence of *o* with bound numerals and prepositions is probably a structural feature related to the fact that the preclitic position is already occupied, with an additional syntactic factor for the preposition as outlined above. Its absence with possessive suffixes is harder to account for.

4. Tolaki

Tolaki is spoken by about 350,000 speakers on the mainland of the southeastern arm of Sulawesi (see Figure 1). It belongs to the Bungku-Tolaki subgroup of Celebic and the data for the following section is taken from Mead & Tambunan (1993), Mead (1994) and Youngman (2001).

The article *o* in Tolaki can be characterised by the following features:

- It has no independent meaning, but for the purposes of this paper is still glossed as ART.
- It only occurs with disyllabic common nouns, never with nouns of three or more syllables. (Monosyllabic nouns do not exist in the language.)
- The grammatical function of the noun phrase containing *o* and its position in the clause are irrelevant parameters for its distribution.
- It is a preclitic morpheme, and usually written as one word with the following noun, a convention which will be followed in this paper.

Examples (18a, b) illustrate *o*= with disyllabic nouns, while (18c, d) show the ungrammaticality of *o* with trisyllabic nouns.

- (18) (a) *o=piso*
ART=knife
'the/a knife'
- (b) *o=donga*
ART=deer
'the/a deer'
- (c) **o=kaluku*
ART=coconut
'the/a coconut'
- (d) **o=laika*
ART=house
'the/a house'

Clausal examples with NPs containing *o=* as subject or object are shown in (19)–(21).

- (19) ... *no-tudu* *o=usa.*
3SG-descend ART=rain
'...it rained.' (Lit. '... rain descended.')
- (20) ... *leu* *o=beka* *sumuko-'i...*
come ART=cat ask-3SG
'...there came a cat and asked him...'
- (21) ... *a-no* *pombiara* *o=bee.*
and-3SG care.for ART=goat
'... and he took care of a goat.'

Within NPs, *o=* is found in combination with demonstratives, some quantifiers and with independent numerals:

- (22) *nggiro'o* *o=piso*
that ART=knife
'that knife'

(23) *dadio* *o=buku*
 many ART=book
 ‘many books’

(24) *oruo* *o=dahu*
 two ART=dog
 ‘two dogs’⁵

In combination with possessive enclitics *o=* does not occur, since the word is already trisyllabic: *o=piso* ‘a/the knife’, but *piso=nggu* ‘my knife’, not **o=piso=nggu*. Similarly: *o=donga* ‘a/the deer’, *donga=no* ‘his/her/its deer’, not **o=donga=no*. However, a disyllabic possessor noun phrase following the possessed NP does take *o=*, as this constitutes a separate phonological word.

(25) *pundi=no* *o=ana* ‘the child’s banana’
 banana=3SG ART=child

(26) *penao=no* *o=donga* ‘the deer’s breath/spirit’
 breath=3SG ART=deer

Notwithstanding these straightforward distributional facts, there are at least four cases when disyllabic common nouns do not take *o=*.

1. As mentioned above, when there is a possessive enclitic, *o=* does not occur, as it violates the rule that the article only attaches to disyllabic nouns.
2. The article is absent when a bound numeral precedes a noun which functions semantically as a measure noun. Compare the following contrastive pair, where (27a) shows the independent numeral with the article simply counting the objects, while in (27b) there is a bound numeral, no article, and the counting does not refer to the objects as such but to their contents.⁶

⁵ Independent numerals also have an initial *o*, but bound numerals do not. Compare the independent *oruo* ‘two’ with bound *rua*, and independent *olimo* ‘five’ with bound *lima*. This morpheme *o* is probably unrelated to the article and remains unglossed.

⁶ In standard Tolaki orthography the bound numerals are written as free words. For the sake of this paper I write them as bound (clitic) morphemes, in line with their phonological status.

(27) (a) *olimo* *o=lepa*
 five ART=basket
 ‘five baskets’

(b) *lima=lepa*
 five=basket
 ‘five basketfuls (of something)’

3. As in Muna, the article is absent when there is a (proclitic) preposition preceding the noun:

(28) *i=wuta*
 LOC=earth
 ‘in/on the earth’

4. The article is absent in certain phrases (possibly phrasal compounds) covering a range of semantic functions, including material composition, generic-specific, association, specification and others. These phrases are characterised by three features: absence of *o=*, the presence of a nasal ligature between the two elements (if the second element starts with a voiceless plosive), and also a tight-knit semantic connection. Some examples are presented in (29).

(29) (a) *ringgi* *wulaa*
 coin gold
 ‘gold coin’

(b) *pu'u* *m-bundi*
 tree LIG-banana
 ‘banana tree’

(c) *usa* *rapo*
 rain dense
 ‘heavy rain’

In conclusion, it appears that the article in Tolaki does not have any syntactic or semantic function. It is merely a prosodic feature associated with disyllabic common nouns. The following prosodic word constraint can be formulated for Tolaki

- (30) Prosodic words which are built on common nouns must be minimally trisyllabic.

The procliticisation of *o=* can then be viewed as a repair mechanism to satisfy this constraint. Notice that the constraint makes reference to both part of speech (common noun) and the number of syllables.

This proposal has several implications for the analysis of the data above:

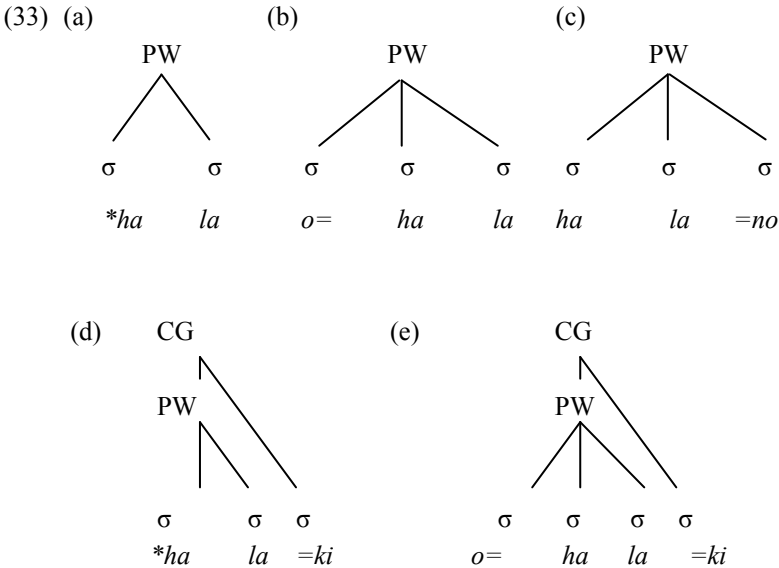
1. Since the article occurs with the nouns in (22)–(24), it follows that demonstratives, some quantifiers and free numerals are all independent phonological words.
2. Possessive enclitics are part of the prosodic word, as shown by examples above following (24). The trisyllabicity constraint is satisfied and no repair mechanism is needed.
3. Prepositions, such as *i=* in (28) are proclitics, thereby creating trisyllabic phonological words.
4. A bound numeral, such as *lima=* in (27b), is also a proclitic and thereby creates a quadrisyllabic prosodic word. In such cases the head noun obtains a specific interpretation as a measure noun.
5. The examples in (29) also constitute polysyllabic single prosodic words, thereby obviating the need for the article. However, there is no cliticisation in these cases and it can be argued that these structures are ‘phrasal compounds’ which consist of two grammatical words united into one prosodic word.

A complication with this analysis is that there are various enclitics following the noun which do co-occur with the article *o=*. Examples include the certainty enclitic *=ki* and the plurality enclitic *=(h)ako*:

- (31) (a) *o=hala* ‘a/the debt’
 (b) *o=hala=ki* ‘certainly a debt’
- (32) (a) *o=sapi* ‘a/the cow’
 (b) *o=sapi=hako* ‘cows’

These facts are somewhat problematic for our analysis, as it would appear that an enclitic such as =*ki* constitutes a single phonological word together with the host noun. However, the fact that the noun still takes the article shows that the enclitic is not counted to be part of the prosodic word. I assume, therefore, that these enclitics are attached to the noun after the prosodic word has been formed. Possibly they are phrasal clitics, but they certainly constitute a separate level in the prosodic hierarchy. Following Nespor & Vogel (1986) and Vogel (2009), I call this the Clitic Group, although this prosodic level is admittedly not universally accepted (see Truckenbrodt 2007).

The following schematic representation in (33) illustrates the contrast on the root *hala* ‘debt’ (*hala=no* ‘his/her debt; *o=hala=ki* ‘certainly a debt’; PW = prosodic word; CG = clitic group). Notice that (a) and (d) are unacceptable as they violate the prosodic word constraint for common nouns.⁷



⁷ A complication occurs with forms such as *sapi=hako=nggu* (cow=PLUR=1SG) ‘my cows’ where the possessive enclitic =*nggu* (part of the prosodic word) follows the plurality clitic =*hako* (part of the clitic group). This is possibly a case of clitic order reversal, the details of which await further investigation.

5. Summary

The following table displays in summary form the essential features of the article in the three languages under discussion.

Table 1: Comparison of the article in Balantak, Muna and Tolaki

		Balantak <i>a</i>	Muna <i>o</i>	Tolaki <i>o=</i>
<i>semantics</i>	independent meaning	no	no	no
<i>syntax</i>	category	common nouns	common nouns	common nouns
	grammatical function of NP	subjects only	any	any
	position in clause	post-predicate	a. clause-initial b. free-standing	any
	with possessive suffixes	yes	no	no
	with bound numerals	yes	no	no
	after prepositions	no	no	no
<i>prosody</i>	shape of noun root	any	any	disyllabic only
	clitic position	enclitic	proclitic	proclitic
	position in intonation phrase	medial	only initial	anywhere

6. Diachrony

Even though the history of the Celebic languages is still poorly understood and reconstruction work is in its infancy, diachronically a number of points can be made regarding these articles.

Firstly, Balantak is the most conservative of these three languages, both phonologically and syntactically. It makes more sense to derive the form and the distribution in Tolaki and Muna from a situation resembling Balantak than the other way round. Given the fact that Proto-Malayo-Polynesian is reconstructed with **a* as a specific common noun marker (Blust 2009), I tentatively reconstruct Proto-Celebic **a* as a subject marker with common nouns.⁸ The article did not occur after prepositions and was almost certainly not subject to prosodic constraints.

In Balantak **a* became an enclitic article, which probably led to its disappearance as a subject marker with clause-initial subjects, as in many cases there would be no preceding host. It retained its function as a marker of subjects, but only in post-predicate position.

In Muna and Tolaki **a* became *o*. This is a common sound change in unstressed syllables in these languages, as shown by, for instance, PMP (Proto-Malayo-Polynesian) **qasiRa* ‘salt’ > Muna *ghohia*, Tolaki *ohio*. In both languages the article also lost its grammatical function, a change which was almost certainly related to the breakdown of the voice system in Celebic. In a further development which is possibly typologically unusual, the article turned into a prosodic feature associated with common nouns. In Muna this association occurred on the prosodic phrase level and acts in conjunction with syntactic structure. In Tolaki the prosodic association happened on the word level, acting as a rescue mechanism to satisfy a prosodic word constraint.⁹ In both cases we are dealing with what might be called ‘morphological junk’, that is, material that is inherited from earlier stages of the language which is no longer functional (Lass 1990). Such a development agrees with what Greenberg (1978) calls the ‘Stage III article’, that is, a former demonstrative or definite article which has simply become a marker of nominality. Both in Tolaki and Muna the article *o* does not appear to have any discernable syntactic or semantic function. It is clearly there, its shape and distribution can

⁸ Mead (1994) makes the claim that the article in Tolaki originated as a demonstrative, but the evidence for this appears to be tenuous.

⁹ Interestingly, there are a number of Tolaki words in which an initial *o*, which was part of the root, has been reanalysed as the article. Examples include *posu* ‘gall bladder’ (PMP **qapeju*, expected **oposu*), *lipa* ‘centipede’ (from a Holle list; PMP **qalipan*; the Mekongga dialect *olipa* has retained the vowel). Data from Mead (1998).

be stated, but with its history as a subject marker it appears to be a prime candidate for the category ‘junk morpheme’.

7. Remaining questions

Though the main contours of the article in these three languages are clear, a number of questions remain and further questions are raised by the preceding analysis:

- What exactly is the relationship between the syntactic and the prosodic constraints on *o* in Muna? Sometimes the two constraints are mutually reinforcing (as with nouns in sentence-initial position or nouns in isolation), but in many cases only one appears to play a role. Does the prosodic constraint always override any syntactic considerations? In other words, can common nouns receive the article in any position, given an adequate prosodic break? What exactly is the role of conjunctions and adverbs? It appears, for instance, that when *dua* ‘also’ follows the verb, a postverbal article is not uncommon.
- What are some possible reasons that the article *o* is absent with possessive suffixes in Muna? Is this related to the fact that the article can be absent with other modifiers such as a following participle or a relative clause? Why is this constraint in the southern Muna dialect less robust? See van den Berg (2004) for details. Could this unusual feature of *o* have been influenced by prosodic constraints at work in the neighbouring language Tolaki?
- Can the Tolaki data shed any light on the discussion of the theoretical status of the Clitic Group?
- What is the role of stress in the prosodic hierarchy, especially in Tolaki?
- What are the articles in other Celebic languages? There is some variety in the forms and probably also in the functions of articles, often with subtle distinctions not yet properly documented. A considerable number of Celebic languages have also lost the article.
- Is there a connection with the article *o* which is found in some West Papuan languages of North-Halmahera (see Holton 2006)? Since the languages are genetically unrelated, this appears to be a case of chance similarity, but there is some evidence for linguistic contact between eastern Sulawesi and North-Halmahera and this question deserves to be explored.

References

- Blust, Robert A. 2009. *The Austronesian languages*. Canberra: Pacific Linguistics.
- Clark, Eve V. 1978. Locationals: existential, locative and possessive constructions. In Joseph H. Greenberg, Charles A. Ferguson & Edith Moravcsik (eds.), *Universals of human language, volume 4: syntax*, 85–126. Stanford University Press.
- Dryer, Matthew S. 2008. Definite articles. Indefinite articles. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The World atlas of language structures online*. Munich: Max Planck Digital Library, chapters 37 and 38. Available online at <http://wals.info/feature/> (8 April 2011).
- Greenberg, Joseph H. 1978. How does a language acquire gender markers? In Joseph H. Greenberg, Charles A. Ferguson & Edith Moravcsik (eds.), *Universals of human language, volume 3: word structure*, 47–82. Stanford: Stanford University Press.
- Holton, Gary. 2006. The relational noun marker in Tobelo (Northeast Halmaheran). *Papers from the Tenth International Conference on Austronesian Linguistics*. www.sil.org/asia/philippines/ical/papers/holton-tobelo.pdf
- Lass, Roger. 1990. How to do things with junk: exaptation in language evolution. *Journal of Linguistics* 26, 79–102.
- Mead, David. 1994. *Was Tolaki prothetic vowel "o" once a demonstrative?* Unpublished typescript, 13 pp. Rice University.
- Mead, David. 1998. *Proto-Bungku-Tolaki: reconstruction of its phonology and aspects of its morphosyntax*. Unpublished PhD dissertation, Rice University.
- Mead, David. 2003. Evidence for a Celebic supergroup. In John Lynch (ed.), *Issues in Austronesian historical phonology*, 115–142. Canberra: Pacific Linguistics.
- Mead, David & Tambunan. 1993. *Tolaki phonology*. Unpublished typescript, 28 pp.
- Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
- Reid, Lawrence A. 2002. Determiners, nouns, or what? Problems in the analysis of some commonly occurring forms in Philippine languages. *Oceanic Linguistics* 41, 296–309.
- Ross, Malcolm. 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In Henry Y. Chang, Lilian M. Huang & Dah-an Ho (eds.), *Streams converging into an ocean. Festschrift in honor of Professor Paul Jen-Kuei Li on his 70th birthday*, 521–563. Taipei: Academia Sinica.
- Truckenbrodt, Hubert. 2007. The syntax-phonology interface. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, 435–456. Cambridge: CUP.
- van den Berg, René. 1989. *A Grammar of the Muna language*. Dordrecht: Foris.

- van den Berg, René. 2004. Notes on the southern Muna dialect. In John Bowden & Nikolaus P. Himmelmann (eds.), *Papers in Austronesian subgrouping and dialectology*, 129–169. Canberra: Pacific Linguistics.
- van den Berg, René & Robert L. Busenitz. to appear. *A grammar of Balantak, a language of Eastern Sulawesi*.
- Vogel, Irene, 2009. The status of the clitic group. In Janet Grijzenhout & Baris Kabak (eds.), *Phonological domains. Universals and deviations*, 15–46. Berlin: Mouton de Gruyter.
- Youngman, Scott. 2001. *Addenda to Tolaki phonology*. Unpublished typescript. 64 pp.